UNITED STATES DEPARTMENT OF AGRICULTURE
Agricultural Marketing Service

MATHEMATICAL AVERAGES

An Office Report

Prepared by
F. H. Harper and Violet M. Feild

Washington, D. C.
November 1, 1939

UNITED STATES DEPARTMENT OF AGRICULTURE
Agricultural Marketing Service

MATHEMATICAL AVERAGES

An Office Report Prepared by
F. H. Harper and Violet M. Feild

- - - - - -

Definition of Arithmetic Mean

The arithmetic mean (average) is an intermediate quantity or value, being of such magnitude that the sum of the plus deviations is the same as the sum of the minus deviations. It may be defined also as the point of central tendency from which the algebraic sum of the deviations is zero. Therefore, if the sum of the plus deviations is subtracted from the sum of the minus deviations the difference will be zero, showing that the arithmetic mean is the point about which the deviations reach a minimum.

In this discussion no differentiation is made between the terms "arithmetic average" and "arithmetic mean," although, as explained by Doctor A. L. Bowley on page 82 of the 1920 edition of his "Elements of Statistics," some writers have attempted to draw a distinction between averages and means, but no general agreement has been reached as to the exact senses in which the words are to be separately applied. In connection therewith Doctor Bowley refers to the article "Moyenne," by Doctor Bertillon, in <u>Dictionnaire encyclopédique des Sciences Médicales</u>, and to the paper by Doctor Venn in the <u>Statistical Journal</u>, 1891, and chapter 18 in his <u>Logic of Chance</u>.

The principle concerning the location of the arithmetic mean is illustrated by table 1, which contains, in the first column, statistics on the amount of new-mortgage loans made by savings and loan associations during the period July 1, 1937, to June 30, 1938, for the purchase of new homes. For the 12-month period the loans in the 12 Federal Home Loan Bank Districts amounted to 286,548 thousands of dollars, averaging 23,879 thousands per district. In each of 6 of the 12 districts the amount of loans was less than the average and in each of the other 6 districts the amount of loans was greater than the average.

The sum of the differences between the average and loans that were less than the average is 66,363 thousands of dollars, which is exactly the same as the sum of the differences between the average and loans that were greater than the average.

Table 1.- Estimated amount of new-mortgage loans made by all savings and loan associations for purchase of homes, July 1, 1937, to June 30, 1938, by districts

Federal Home Loan Bank District 1/	Loans made for purchase of homes 2/	Deviation from the mean	
		Minus	Plus
	1,000 dollars	1,000 dollars	1,000 dollars
1	32,650	---	8,771
2	31,136	---	7,257
3	30,571	---	6,692
4	28,761	---	4,882
5	58,261	---	34,382
6	12,785	11,094	---
7	28,258	---	4,379
8	14,830	9,049	---
9	14,114	9,765	---
10	14,472	9,407	---
11	6,308	17,571	---
12	14,402	9,477	---
Total	286,548	66,363	66,363
Mean	23,879	---	---

1/ Boundaries of districts are shown in the map on page 20 of the Sixth Annual Report of the Federal Home Loan Bank Board, for the period July 1, 1937--June 30, 1938.

2/ Sixth Annual Report of the Federal Home Loan Bank Board, for the period July 1, 1937--June 30, 1938, exhibit 8, page 105.

Although the arithmetic mean is generally considered as the value that is of such magnitude that the algebraic sum of the deviations from it is zero, means as actually expressed and used are frequently not of such magnitudes that sums of the plus and minus deviations are equal. This is because of decimals that are dropped from the quotient in expressing the means, as illustrated by table 2, which contains, in the first column, statistics on the production of gold in the United States during the 16-year period 1922-37.

Table 2.- Production of gold in the United States, 1922-37

Calendar year	Production of gold 1/	Deviation from the mean	
		Minus	Plus
	Fine ounces	Fine ounces	Fine ounces
1922	2,363,075	407,522	---
1923	2,502,632	267,965	---
1924	2,528,900	241,697	---
1925	2,411,987	358,610	---
1926	2,335,042	435,555	---
1927	2,197,125	573,472	---
1928	2,233,251	537,346	---
1929	2,208,386	562,211	---
1930	2,285,603	484,994	---
1931	2,395,878	374,719	---
1932	2,449,032	321,565	---
1933	2,556,246	214,351	---
1934	3,091,183	---	320,586
1935	3,609,283	---	838,686
1936	4,357,394	---	1,586,797
1937	4,804,540	---	2,033,943
Total	44,329,557	2/ 4,780,007	2/ 4,780,012
Mean	2,770,597	---	---

1/ U. S. Treasury Department. Bureau of the Mint. Annual Report of the Director of the Mint, Fiscal Year ended June 30, 1938, page 33. The production in terms of fine ounces is the production of pure gold.

2/ If the mean were expressed as 2,770,597.3125, the quotient obtained by dividing 44,329,557 by 16, the sum of these deviations would be 4,780,010.75. The sum of the minus deviations shown, 4,780,007, would be increased by 3.75 (the product of 12 and .3125) and the sum of the plus deviations shown, 4,780,012, would be decreased by 1.25 (the product of 4 and .3125).

The expressed mean is 2,770,597, but the actual mean is 2,770,597.3125, as explained in footnote 2 of the table. Because of the dropping of decimals in the calculation of the mean the sum of the minus deviations is 5 less than the sum of the plus deviations. If the mean were expressed as 2,770,597.3125 the sum of the minus deviations, 4,780,007, would be increased by an amount equal to the product of .3125 and 12, there being 12 years for which the reported production is less than the mean. That being the case, each of the 12 minus deviations in table 2 would be increased by .3125if the average were expressed as 2,770,597.3125, so that the sum of the minus deviations would be 4,780,007 plus 3.75 (the product of .3125 and 12,) or 4,780,010.75.

For the same reason that the sum of the minus deviations in table 2 is less than it would be if the mean were expressed as 2,770,597.3125 instead of as 2,770,597 the sum of the plus deviations is greater than it would be if the mean were expressed as 2,770,597.3125. This is because the production of each of the years 1934-37, inclusive, exceeds the expressed average, 2,770,597, to a greater extent than the more refined average, 2,770,597.3125. If the latter were expressed as the average the sum of the plus deviations, 4,780,012, would be decreased by an amount equal to the product of .3125 and 4, there being, as indicated, 4 years for which the reported production is greater than the mean. Since this is true, each of the 4 plus deviations in table 2 would be decreased by .3125 if the average were expressed as 2,770,597.3125.

By increasing the sum of the minus deviations shown in table 2 by the product of .3125 a total of 4,780,010.75 is obtained, as we have already observed, and a reduction in the sum of the plus deviations shown in the table by an amount equal to the product of .3125 and 4 leaves 4,780,010.75. The true sum of each series of deviations, therefore, is 4,780,010.75, instead of 4,780,007 and 4,780,012, respectively, which are shown in table 2.

The illustration afforded by table 2 will indicate, as it is intended to, that in many instances the values that are designated as means are actually only the approximate means. This is true, but it is also true that means which are approximate only in the sense that the quotient has been rounded are generally sufficiently precise for the purposes intended. In some instances, of course, it may be desirable to express the mean in more decimals than in others, depending in part upon the magnitudes of items being averaged, the purpose of the averages, and the preferences of those making the calculations.

The arithmetic mean is one of the most, if not the most, commonly used of all statistical measures, and it is probably more readily comprehended than any other. The ease with which the arithmetic mean is comprehended is probably attributable to the ease with which it is calculated.

Simple Arithmetic Mean

The simple arithmetic mean is the quotient obtained by dividing a sum by the number of items of which it is composed. Table 3, containing statistics on the number of passengers carried on air-mail routes during the 12-month period that ended with June, 1938,

illustrates its calculation. According to the table, 1,234,696 passengers were carried during the period, an average of 102,891 passengers per month.

The mean was obtained by summating the monthly statistics to obtain the total for the 12-month period and then dividing by 12. The formula for this procedure is $\frac{\Sigma X}{N}$, in which the meanings of the symbolic expressions are as follows:

ΣX, the total number of passengers carried
N, the number of months.

Another formula, $\frac{1}{N}\Sigma X$, may be used in calculating the mean. By using this formula the total is multiplied by the quotient obtained by dividing the number of items into 1. There are 12 items in table 3. The quotient obtained by dividing 1 by 12 is .083333, and the product of .083333 and 1,234,696 is 102,891, which is the monthly average number of passengers carried.

Table 3.- Passengers carried on air-mail routes, by months, fiscal year ended June, 1938

Month	Passengers carried 1/
	Number
July	119,257
August	123,754
September	125,520
October	112,991
November	83,511
December	71,969
January	73,642
February	78,153
March	98,916
April	107,854
May	120,738
June	118,391
Total	1,234,696
Mean	102,891

1/ Annual Report of the Postmaster General, Fiscal Year ended June 30, 1938, table 51, page 132.

Although the arithmetic average is useful in statistical analyses, it furnishes no indication of the range in magnitude of items. The averaging process in effect smooths out the variations and gives a value which when multiplied by the number of items recovers the original sum of the items, if decimals are carried to a sufficient number of places.

Table 4 contains statistics on the number of clear days reported by the Weather Bureau for 3 cities in the United States. As will be observed, the total number of clear days reported for the 12 months is the same in each instance and, consequently, the monthly average is the same. However, there is considerable difference in the ranges within the 12-month period. For Tampa the number of clear days ranged from 5 in July and August to 14 in March; for Baltimore the number of clear days ranged from 9 in January, February, and December to 13 in October; and for Nashville the number of clear days ranged from 8 in January and December to 15 in October. Neither the total number of clear days nor the monthly average for the 3 cities furnishes any indication of the difference in range, which is clearly shown by the table. It is for this reason that totals and averages might be misleading if they are shown without the individual items from which they are derived.

Table 4.- Long-time average number of clear days at Tampa, Florida; Baltimore, Maryland; and Nashville, Tennessee, by months 1/

Month	Clear days at Tampa, Florida 2/	Clear days at Baltimore, Maryland 2/	Clear days at Nashville, Tennessee 2/
	Number	Number	Number
January	11	9	8
February	11	9	7
March	14	10	9
April	13	10	9
May	11	10	10
June	7	9	9
July	5	10	10
August	5	10	12
September	7	12	13
October	12	13	15
November	13	10	11
December	12	9	8
Annual	121	121	121
Mean	10	10	10

1/ U. S. Department of Commerce. Bureau of Foreign and Domestic Commerce. Statistical Abstract of the United States, 1937, table 144, pages 135, 137, and 142 (from Weather Bureau, U. S. Department of Agriculture.)

2/ Long-time averages.

Table 5 contains statistics on the harvested acreage and on the production of lima beans in the United States for the 14-year period 1924-37. During this period the acreage ranged from 2,000 acres in 1924, 1925, and 1926 to 13,120 acres in 1932, and production ranged from 160,000 bushels in 1926 to 908,000 bushels in 1932. The averages, 8,071 acres and 504,000 bushels, furnish no indication of these ranges, however, indicating in some measure the inadequacy of averages when considered alone in the study of variation.

Table 5.- Acreage and production of lima beans in the United States for marketing in the fresh state, 1924-37 1/

Year	Acreage harvested	Production
	Acres	Bushels
1924	2,000	172,000
1925	2,000	200,000
1926	2,000	160,000
1927	3,530	285,000
1928	5,170	270,000
1929	5,050	386,000
1930	10,960	649,000
1931	11,870	790,000
1932	13,120	908,000
1933	11,850	568,000
1934	11,850	548,000
1935	9,500	567,000
1936	11,400	863,000
1937	12,700	689,000
Mean	8,071	504,000

1/ U. S. Department of Agriculture. Agricultural Statistics, 1938, table 198, page 156. The Baltimore Evening Sun of Thursday, January 19, 1939, in commenting on the origin of lima beans explained that "The Lima Bean gets its name from Lima, Peru, where a naval officer got some seed and brought them to the United States in 1824."

The simple arithmetic mean can be calculated by assuming a mean, determining the deviations of the individual items from the assumed mean, summating these deviations algebraically, dividing by the total number of deviations, and then adding the quotient to the assumed mean or subtracting it from the assumed mean, depending upon the sign carried by the algebraic sum of the deviations. Table 6 will clarify this procedure.

The total number of officers assigned was 34,602. In calculating the mean, or average, number per corps area the mean was assumed to be 4,000. Deviations of the number of officers assigned to units in individual corps areas were then determined and recorded in the column headed "Deviations from the assumed mean," the minus deviations being shown in one column and the plus deviations in another.

The difference between the sum of the minus deviations and the sum of the plus deviations is 1,398. Since this is a minus quantity, the sum of the minus deviations being greater than the sum of the plus deviations, the quotient obtained by dividing 1,398 by 9, the total number of deviations, will be subtracted from the assumed mean to obtain the true mean. The quotient obtained by division is 155, and when this is subtracted from 4,000 the remainder is 3,845, the mean. If the algebraic sum of the deviations from the assumed mean had been a plus quantity the quotient resulting from dividing the algebraic sum of the deviations by the number of deviations would have been added to the assumed mean to obtain the true, or actual, mean.

Table 6.- Officers assigned or attached to divisional units of Organized Reserves, by corps areas, fiscal year ended June 30, 1938

Corps area	Officers assigned or attached 1/	Deviation from the assumed mean	
	Number	Minus	Plus
First	2,828	1,172	---
Second	3,765	235	---
Third	3,886	114	---
Fourth	5,253	---	1,253
Fifth	2,821	1,179	---
Sixth	3,741	259	---
Seventh	3,420	580	---
Eighth	4,328	---	328
Ninth	4,560	---	560
Total	34,602	3,539	2,141
Assumed Mean	4,000	---	---
True Mean	3,845	---	---

1/ Annual Report of the Secretary of War, 1938, page 68.

Whenever an assumed mean is greater than the true mean the algebraic sum of deviations from it will be a minus quantity, showing that the assumed mean must be reduced in magnitude. Whenever an assumed mean is less than the true mean the algebraic sum of deviations therefrom will be a plus quantity, showing that the assumed mean must be increased. If calculating machines are available it is generally preferable to consider zero as the assumed mean, so that in calculating the arithmetic mean it is only necessary to summate the items and divide by the number of items, the items, themselves, being plus deviations from the assumed mean. In effect, therefore, when an arithmetic mean is calculated by summating the items as they occur in the series and then dividing by the number of items the mean is assumed to be zero, and the items, themselves, are plus deviations from the assumed mean.

The nature of differences between arithmetic means calculated by 2 methods is indicated by tables 7 and 8. Table 7 contains statistics for 1936 on the acreage and production of corn in the North Central States and on the mean yield per acre, by States. The mean yield for each individual State was calculated by dividing production by acreage harvested, and the mean yield for the North Central States as a whole was calculated by dividing total production by total acreage harvested.

Table 8 shows the mean yields for individual States and the mean for the group as a whole. The latter mean, 15.5, was obtained by summating the means for individual States and dividing the sum

by 12, the number of means summated. By this method of calculation the mean for each State has equal weight, regardless of acreage and producton. The mean yield in Kansas, for example, of 4.0 bushels has as much weight as the mean yield of 33.0 bushels in Ohio, although, as shown by table 7, the acreage harvested in Kansas was only 2,759,000 acres, compared with 3,685,000 acres harvested in Ohio, and although the production in Kansas was only 11,036,000 bushels, compared with a production of 121,605,000 bushels in Ohio.

Obviously, the mean of 15.5 bushels shown in table 8 is not the true average yield, then, for the North Central States as a whole because no account has been taken in its calculation of differences in acreage and production reported for the States comprising the group. If the acreage and production were the same for every State, then the mean of the sum of the State means would be the same as the mean obtainable by dividing total production in the group of States by total acreage, but under no other circumstances would the 2 means be alike, except by coincidence.

Table 7.- Acreage, production, and average yield of corn per acre in the North Central States, 1936 1/

State	Acreage harvested	Total production	Average yield per acre
	1,000 acres	1,000 bushels	Bushels
Ohio	3,685	121,605	33.0
Indiana	4,569	116,510	25.5
Illinois	9,266	217,751	23.5
Michigan	1,500	36,750	24.5
Wisconsin	2,204	44,080	20.0
Minnesota	4,649	88,331	19.0
Iowa	10,759	190,434	17.7
Missouri	5,004	40,032	8.0
North Dakota	744	2,530	3.4
South Dakota	2,484	8,446	3.4
Nebraska	7,674	26,859	3.5
Kansas	2,759	11,036	4.0
All States	55,297	904,364	2/ 16.4

1/ U. S. Department of Agriculture. Agricultural Statistics, 1938, table 43, page 44.

2/ Calculated by dividing total production by total acreage harvested.

Table 8.- Average yield of corn per acre in the North Central States, 1936

State	Average yield per acre 1/
	Bushels
Ohio	33.0
Indiana	25.5
Illinois	23.5
Michigan	24.5
Wisconsin	20.0
Minnesota	19.0
Iowa	17.7
Missouri	8.0
North Dakota	3.4
South Dakota	3.4
Nebraska	3.5
Kansas	4.0
All States	15.5

1/ U. S. Department of Agriculture. Agricultural Statistics, 1938, table 43, page 44.

2/ Calculated by dividing the sum of the average yields by the number of States. This is not the true average. See table 7.

Arithmetic Mean of a Frequency Distribution

When observations are arranged in the form of a frequency distribution the arithmetic mean can be calculated readily by multiplying the frequencies by the classes, by the midpoints of the classes, or by some other point within the classes, as the case may be; summating the products thus obtained; and dividing by the number of frequencies.

Table 9 shows the grades received by 214 college students on a test in economic theory. These grades ranged from 89 to 100. The average of the 214 grades is 94, expressed as a whole grade, calculated by dividing the sum of the products of frequencies and grades, 20,208, by the number of grades, 214, the sum of the products of frequencies and grades being, in effect, the sum of the 214 individual grades.

Table 9.- Grades received by 214 college students on a test in economic theory 1/

Grade received (class)	1 Number of students to whom test was given (frequency)	2 Product of frequency and grade received
89	8	712
90	10	900
91	16	1456
92	18	1656
93	25	2325
94	38	3572
95	26	2470
96	20	1920
97	19	1843
98	17	1666
99	12	1188
100	5	500
Total	214	20208

1/ Specially selected for illustrative purposes. The mean of the 214 grades is 94.

Table 9 is a frequency distribution of the simplest kind. There is a class interval of 1, as shown by the successive magnitudes of the grades in the stub, and the frequencies in column 1 opposite the grades with which they are associated show clearly the number of students who received each grade. The products of the frequencies and grade in column 2, as already indicated, are equal to the sums of the grades. For example, the product 712 is equal to the sum of the 8 grades of 89.

Distributions of a somewhat different arrangement are shown in tables 10 and 11. So far as calculation of the average cost per ton for all 5 areas as a whole is concerned, only the information in the last row of columns 2 and 3 is necessary, but the other information might be of interest in interpreting the statistics on total production, total cost, and average cost per ton.

In both of these tables the mean cost per ton for each area was determined from the information in columns 2 and 3, as was also the mean cost per ton in the 5 areas as a whole. It will be observed that the arrangement of the items in columns 2 and 3 of table 10 is the reverse of the arrangement in table 11, which necessarily reverses the arrangement of items in column 4. However, the average cost per ton for the 5 areas as a whole is the same in both instances, although the average costs per ton shown in table 10 for individual areas are different from average costs shown in table 11 for the same areas.

Table 10.- Production and costs of production of rutabagas on 150 farms, 1937, by areas 1/

Area in which farms were located	1	2	3	4
	Farms represented	Total production	Total cost of production	Mean cost of production per ton
	Number	Tons	Dollars	Dollars
1	10	100	50	0.50
2	20	200	200	1.00
3	30	300	450	1.50
4	40	400	800	2.00
5	50	500	1,250	2.50
All areas	150	1,500	2,750	2/ 1.83

1/ Specially selected for illustrative purposes.
2/ Calculated by dividing total cost, $2,750, by total production, 1,500 tons.

Table 11.- Production and costs of production of rutabagas on 150 farms, 1938, by areas 1/

Area in which farms were located	1	2	3	4
	Farms represented	Total production	Total cost of production	Average cost of production per ton
	Number	Tons	Dollars	Dollars
1	10	500	1,250	2.50
2	20	400	800	2.00
3	30	300	450	1.50
4	40	200	200	1.00
5	50	100	50	.50
All areas	150	1,500	2,750	2/ 1.83

1/ Specially selected for illustrative purposes.
2/ Calculated by dividing total cost, $2,750, by total production, 1,500 tons.

Tables 12 and 13 contain frequency distributions of which the averages were calculated by multiplying the midpoints of the classes by the number of frequencies, summating the products thus obtained, and then dividing by the total number of frequencies.

Table 12.- Staple length of upland cotton ginned in the United States, crop of 1935

Staple length (inches)	1 Ginnings in running bales 1/ (frequency)	2 Product of number of bales and midpoints of staple-length groups 2/
Shorter than 7/8	1,320,100	17,821,350
7/8 and 29/32	3,235,100	46,908,950
15/16 and 31/32	2,628,100	40,735,550
1 and 1-1/32	1,682,200	27,756,300
1-1/16 and 1-3/32	866,500	15,163,750
1-1/8 and 1-5/32	554,000	10,249,000
1-3/16 and 1-7/32	102,600	2,000,700
1-1/4 and longer	14,100	289,050
Total	3/ 10,402,700	160,924,650
Mean 4/ 15.47		

1/ U. S. Department of Agriculture. Agricultural Statistics, 1938, table 124, page 105.

2/ For purposes of the calculations, the midpoints of the staple-length groups were assumed to be as follows, in sixteenths of an inch: 13.5, 14.5, 15.5, 16.5, 17.5, 18.5, 19.5, and 20.5.

3/ According to the Bureau of the Census, total ginnings from the United States crop of 1935 amounted to 10,420,346 running bales, of which, 17,619 bales were American-Egyptian cotton. The ginnings of 175 running bales of Sea-Island cotton are included in the Department's staple-length statistics for upland cotton.

4/ Sixteenths of an inch, calculated by dividing the sum of column 2 by the sum of column 1. See footnote 2.

Table 13.- Frequency distribution of farm costs of producing 2,280 tons of alfalfa hay 1/

Cost in dollars per ton (class)	1	2
	Tons produced (frequency)	Product of tons produced and midpoint of class 2/
1.51 - 2.00	10	17.50
2.01 - 2.50	60	135.00
2.51 - 3.00	330	907.50
3.01 - 3.50	370	1,202.50
3.51 - 4.00	330	1,237.50
4.01 - 4.50	340	1,445.00
4.51 - 5.00	300	1,425.00
5.01 - 5.50	130	682.50
5.51 - 6.00	160	920.00
6.01 - 6.50	90	562.50
6.51 - 7.00	30	202.50
7.01 - 7.50	50	362.50
7.51 - 8.00	60	465.00
8.01 - 8.50	20	165.00
	2,280	9,730.00
Mean cost per ton:	---	4.27

1/ Specially selected for illustrative purposes.

2/ Costs, in dollars, used as midpoints of the classes are as follows: 1.75, 2.25, 2.75, 3.25, 3.75, 4.25, 4.75, 5.25, 5.75, 6.25, 6.75, 7.25, 7.75 and 8.25.

The mean of a frequency distribution can be calculated by assuming a mean, determining the deviations therefrom, multiplying the deviations by the frequencies, summating algebraically, dividing the algebraic sum by the number of deviations, and then adding the quotient to the assumed mean or subtracting it from the assumed mean, depending upon the sign carried by the algebraic sum. Table 14 illustrates the calculation of the arithmetic mean of a frequency distribution by the use of an assumed mean. The assumed mean of the distribution of 16 grades is 92, whereas the true mean is 93.

The true mean was obtained by multiplying the frequencies by the grades, summating the products, and dividing by the number of grades. The sum of the products is 1488, and the total number of grades is 16, the same as the number of students who took the test. The quotient obtained by dividing 1488 by 16 is 93, the average of the 16 grades. In this instance, of course, one can readily see that the mean is 93 because there are as many grades less than 93 as there are grades that are greater, but the table serves as well to illustrate the procedure described as though the true mean were not so readily discernible.

By assuming a mean the true mean can be obtained in the manner indicated. The sum of the minus deviations from the assumed mean is 4, whereas the sum of the plus deviations is 20, so that the algebraic sum of the deviations is 16, the difference between 20 and 4. The quotient obtained by dividing this algebraic sum by the number of deviations is 1. To obtain the true mean, 1 is added to 92, the assumed mean, because the algebraic sum of the deviations is a plus quantity.

Table 14.- Grades received by 16 college students on a marketing methods test 1/

Grade received (class)	Students to whom test was given (frequency)	Deviation of grade from the assumed mean of grades (assumed mean is 92)		Product of frequency and deviation of grade from the assumed mean of grades	
	Number	Minus	Plus	Minus	Plus
90	1	2	---	2	---
91	2	1	---	2	---
92	3	---	---	---	---
93	4	---	1	---	4
94	3	---	2	---	6
95	2	---	3	---	6
96	1	---	4	---	4
Total	16	---	---	4	20

1/ Specially selected for illustrative purposes. The true mean is obtainable by adding to 92, the assumed mean, the quotient, 1, obtained by dividing 16 into the difference between the sums of the minus and plus deviations. If this difference had been a minus quantity the quotient would have been subtracted from the assumed mean to obtain the true mean.

Another calculation of the arithmetic mean from an assumed mean is illustrated by table 15. The true mean [1] of costs per ton is \$2.80, obtained by dividing the sum of the products of the frequencies and the midpoints of classes by 200, the total number of tons produced. The true mean is obtained also by the process of assuming a mean, determining the deviations of the midpoints of the classes from the assumed mean, multiplying these deviations by the corresponding frequencies, summating the products algebraically, dividing the algebraic sum by the total number of deviations, and then subtracting the quotient from the assumed mean. The quotient is subtracted in this instance from the assumed mean because the algebraic sum of the deviations is a minus quantity, which shows that the assumed mean is greater than the true mean. If the assumed mean were less than the true mean the algebraic sum of the deviations therefrom would be a plus quantity, in which instance the quotient obtained by dividing this sum by the total number of deviations would be added to the assumed mean to obtain the true mean.

1/ That is, the true mean insofar as it is obtainable from information contained in the table. Obviously, the mean might have a more precise meaning if the sum of the 200 individual costs were divided by the total number of tons produced, but the individual costs per tons are not shown.

Table 15.- Frequency distribution of costs of production of 200 tons of clover hay 1/

Cost in dollars per ton (class)	Tons produced (frequency)	Deviation of midpoint of class from the assumed mean (assumed mean is \$3.00)		Product of frequency and deviation of midpoint of class from the assumed mean	
	Number	Minus	Plus	Minus	Plus
1.51 - 2.00	40	1.25	---	50.00	---
2.01 - 2.50	30	.75	---	22.50	---
2.51 - 3.00	60	.25	---	15.00	---
3.01 - 3.50	30	---	.25	---	7.50
3.51 - 4.00	20	---	.75	---	15.00
4.01 - 4.50	20	---	1.25	---	25.00
Total	200	---	---	87.50	47.50

1/ Specially selected for illustrative purposes. The desired mean is obtainable by subtracting from \$3.00, the assumed mean, the quotient, \$0.20, obtained by dividing 200 into \$40.00, the difference between the sums of the minus and plus deviations. If this difference had been a plus quantity the quotient would have been added to the assumed mean to obtain the desired mean.

The preceding illustrations will show that to calculate the mean of frequency distributions of continuous data of the nature of those described it is only necessary to multiply the midpoints of the classes by the number of frequencies, summate the products, and divide by the total number of frequencies. If the data are of the discrete type the classes, or midpoints of the classes, as the case may be, are also multiplied by the frequencies, the products summated, and the sum divided by the total number of frequencies.

In the case of either discrete or continuous data the assumption is that the midpoints of the classes, when they are used, are the means of the magnitudes of frequencies in the classes or that the extent to which the midpoints in some instances are greater than the means of the classes is offset by other instances in which the midpoints of the classes are less than the means. This assumption is generally permissible when there are large numbers of frequencies. If the midpoints of the classes represent the mean magnitudes of the frequencies in the classes it is obvious that the algebraic sum of deviations from the midpoints is zero. That is, for example, if in table 15 $1.75 is the average cost of producing the 40 tons of hay in the $1.51 - $2.00 class, then the algebraic sum of the deviations of individual costs from $1.75 would be zero.

Weighted Arithmetic Mean

There may be instances in which it is desired to apply weights to a series of means to obtain a mean that describes the series as a whole. For example, if we were given statistics on the acreage of corn harvested in each of the North Central States and on the mean yield per acre and were asked to calculate a mean yield for the group of States

as a whole it would be necessary, assuming that statistics on production were not available, (1) to obtain an estimate of production in each State by multiplying acreage by mean yield per acre, summate the products, and divide the sum by total acreage; (2) apply weights to the mean yields reported for individual States, summate the products, and divide the sum by the sum of the weights; or (3) summate the mean yields reported for the States and divide by the number of State mean yields.

Obviously, the quotient that would be obtained by dividing the sum of State means by the number of means would be only a mean of a series of means. It would not likely be the desired mean because each individual mean would carry equal weight regardless of differences in acreage and production from one State to another. The mean yield for a State in which 1,000 acres were harvested, for example, would carry as much weight as the mean yield for a State in which 10,000 acres were harvested. No account would be taken at all of acreage differences by this procedure.

In order for the desired mean yield for the group of States as a whole to be obtained it would be necessary to take into consideration the differences in acreage harvested from one State to another. Table 16 is shown to illustrate a procedure that can be resorted to in calculating an average yield of corn per acre in a group of States as a whole when only statistics on yields and acres harvested are available for individual States.

In estimating expectancy, however, the simple arithmetic mean of a series of means may have some value.

Table 16.- Acreage and average yield of corn per acre in the North Central States, 1936 1/

State	Acreage harvested (1,000 acres)	Average yield per acre (bushels)	Weight (Kansas yield as the base)	Product of weight and average yield per acre
Ohio	3,685	33.0	1.3356	44.07480
Indiana	4,569	25.5	1.6560	42.22800
Illinois	9,266	23.5	3.3585	78.92475
Michigan	1,500	24.5	.5437	13.32065
Wisconsin	2,204	20.0	.7988	15.97600
Minnesota	4,649	19.0	1.6850	32.01500
Iowa	10,759	17.7	3.8996	69.02292
Missouri	5,004	8.0	1.8137	14.50960
North Dakota	744	3.4	.2697	.91698
South Dakota	2,484	3.4	.9003	3.06102
Nebraska	7,674	3.5	2.7814	9.73490
Kansas	2,759	4.0	1.0000	4.00000
All States	55,297	16.4	20.0423	327.78462

1/ U. S. Department of Agriculture. Agricultural Statistics, 1938, table 43, page 44. The average yield per acre for the group of States as a whole is obtainable by dividing 327.78462 by 20.0423.

The weights in the 3rd column were calculated by dividing the number of acres reported for Kansas into the number of acres reported for each individual State. This gives a weight of 1.0000 for Kansas, the weights for other States varying from the weight for Kansas according to differences between acreages harvested in them and in Kansas. The acreage for any other State, of course, could be taken as the base, with no different effect on final results.

In the table, as will be observed, the yields per acre have been multiplied by the weights and the products recorded in the last column. When the sum of these products is divided by the sum of the weights a mean yield of 16.4 bushels is obtained for the North Central group of States as a whole, which is the same as the mean shown in the last row of the 3rd column of table 7, calculated by dividing total production by total acreage harvested.

In calculating the mean of the yields in the 2nd column of table 16 by the procedure described the numbers of acres, which are the real frequencies, were used in obtaining the weights, and in calculating the mean cost of $4.27 shown in table 13 all values of the variable, cost per ton, except the midpoints of the classes, were ignored. In the latter instance the calculation, in effect, was the determination of the mean of the midpoint values listed in footnote 2 of the table, each midpoint being weighted by the number of frequencies in the classes, and in the calculation of the mean of all the grades in table 9 each individual grade listed in the stub was weighted by the number of grades of each magnitude, just as the

midpoints of the staple-length groups in the stub of table 12 were weighted by the number of bales of cotton reported for each length group. However, in the calculation of the mean of a frequency distribution the actual fact is that the mean obtained is merely the quotient resulting from dividing a sum which actually is, or which is assumed to be, in effect, the equivalent of the total of all the magnitudes, by the total number of frequencies. Therefore, the mean of a frequency distribution of the nature described is a weighted mean only in the sense that a magnitude for each frequency is included in the sum that is divided by the total number of frequencies, so that, as in the calculation of the mean of an array, each item is counted 1 time only. This is as it should be.

The average, 16.4, shown at the bottom of the 2nd column in table 16 can be calculated also by dividing the total acreage, 55,297, into the acreage for each State to obtain a series of weights, multiplying the yields in the 2nd column by the corresponding weights, summating the products, and dividing by the sum of the weights. The sum of these weights, following, is 1.

Ohio	.06664
Indiana	.08263
Illinois	.16757
Michigan	.02713
Wisconsin	.03986
Minnesota	.08407
Iowa	.19457
Missouri	.09049
North Dakota	.01345
South Dakota	.04492
Nebraska	.13878
Kansas	.04989
	1.00000

When the yields in the 2nd column of table 16 are multiplied by these weights, the products summated, and the sum divided by 1 an average of 16.4 bushels per acre is obtained for the group of States as a whole.

In comparatively recent years there has been developing an ever-increasing tendency toward discontinuing the use of the term "weighted mean" in referring to the arithmetic mean of a frequency distribution and to use the term in a more limited sense. This is somewhat of a change in terminology that was very common no longer ago than 10 years. Analysts are realizing that since a frequency distribution is merely an array reduced to more compact form the mean thereof belongs in the same category as the mean of an array.

The arithmetic mean yield of corn per acre calculated by the procedure described in connection with table 16 for the North Central group of States as a whole is weighted in the sense that each individual State's mean yield is included in the sum that was divided by the sum of the weights in proportion to the relation between the number of acres associated with each State's yield and the number of acres associated with the mean that is given a weight of 1. The number of acres could have been used as the weights, with the same result as that accomplished by use of the calculated weights in the 3rd column of table 16. By use of either the calculated weights or the acreage figures, themselves, as weights the final result is the same because each mean is counted in proportion to acreage, in one instance, as indicated, the mean yield for each State being included

as many times as there are number of acres reported, and in the other instance the mean yield for each State being included as many times as the number of acres associated therewith is greater than the number of acres associated with the mean that is given a weight of 1, or a proportion of the mean yield for each State being included that is equal to the proportion that the acreage for the State is of the acreage associated with the mean that is given a weight of 1.

Present usage of the term "weighted mean" restricts it somewhat to means obtained by the use of weights of the general nature of those listed in the 3rd column of table 16. Although there are reasons for not referring to the mean (16.4) in the last row of the 2nd column of that table as a weighted mean, some analysts are inclined to consider it a weighted mean to the same extent that a consumer's index of cost of living is a weighted mean index.

In the calculation of the mean shown at the bottom of the 2nd column of table 16 for the North Central States as a whole no difficulty was experienced in obtaining a mean of the same magnitude as the mean shown at the bottom of the 3rd column of table 7 because the acreage figures were available to use in calculating a series of weights. By using the acreage figures in calculating the weights it necessarily follows that the mean at the bottom of the 2nd column of table 16 must be of the same magnitude as the mean at the bottom of the 3rd column of table 7.

Let us assume, though, that with only the means available that are shown in the 2nd column of table 16 it is desired to calculate a

mean yield for the North Central States by attempting to assign weights to means of individual States in proportion to their relative importance. The assigning of weights for use in applying to the individual items would result in a mean that some analysts might describe as a weighted mean. Only by coincidence, or only because of knowledge possessed by the analyst concerning precise differences in acreage from one State to another, would the weights assigned to the means result in a mean for the group of States as a whole of the same magnitude as the mean shown at the bottom of the 2nd column of table 16 and at the bottom of the 3rd column of table 7, but the mean so calculated might be considered by some analysts a weighted mean in the same sense that a consumer's index of cost of living is a weighted mean index.

The applicability of the term "weighted mean" to a consumer's index of cost of living can be illustrated briefly. Let us assume that for a certain year the following were the indices of cost of living:

Food	110
Clothing	109
Rent	100
Fuel	94
Light	98
Miscellaneous items ...	100

Let us assume further that these costs represented the following percentages of the total cost-of-living expenditures:

Food	38	percent
Clothing	18	"
Rent	18	"
Fuel	4	"
Light	2	"
Miscellaneous items	20	"
	100	"

The consumer's index would be calculated by multiplying the indices by the percentages, which will be used as weights, summating the products, and dividing the sum of the products by the sum of the weights. The calculations are as follows:

110	times	38	equals	4,180
109	times	18	equals	1,962
100	times	18	equals	1,800
94	times	4	equals	376
98	times	2	equals	196
100	times	20	equals	2,000
		100		10,514

The mean index is 105.14, obtained by dividing the sum of the products, 10,514, by the sum of the weights, 100. A mean thus obtained as a measure of cost of living is commonly referred to by statisticians as a weighted mean. Such means are frequently calculated in the construction of index numbers, since, for instance, the index of food costs in general actually may be a weighted average of the indices of costs of various items of food, and the index of

clothing costs may be a weighted average of the indices of costs of the different articles of clothing.

One of the best illustrations, perhaps, of a weighted mean is one that is calculated to show the average grade made by students in a given course when, let us say, a weight of 5 is given to oral reports on supplementary reading; 6, to classroom work; 4, to periodic examinations; 3, to the final examination; and 2, to written assignments. In this instance a weighted mean could be calculated by multiplying the grades for each phase of the work by the corresponding weights, summating the products, and then dividing by the sum of the weights, which is 20.

Progressive and Cumulative Mean

In some instances a series of means calculated by including an additional item every time a mean is to be calculated are useful in studying variability. Such a series of means consists of means of cumulations, or of progressive summations, as some may prefer to describe such cumulated totals. The authors have used the term "progressive mean" for many years, although it has been realized that the term is sometimes applied to a moving average that is calculated by weighting items near the center of the groups more heavily than other items in the groups arranged for calculating the series of moving averages.

Although there are differences of opinion concerning the applicability of the terms "progressive means" and "cumulative means" to the means calculated by successively including an additional item in the total every time a mean is to be calculated, there are some

eminent analysts who consider that the terms appropriately describe a series of means so calculated. Among the younger present-day analysts who hold this view are A. Mason DuPrè and R. F. Hale, the superior judgment of both of whom is beyond question so far as the authors are concerned.

The following calculations illustrate the procedure for determining the so-called progressive, or cumulative, means, as the terms are herein described.

Item	Sum	Mean
1	-	-
2	3	1.5
3	6	2.0
4	10	2.5
5	15	3.0
6	21	3.5
7	28	4.0
8	36	4.5
9	45	5.0

Harmonic Mean

The harmonic mean (average) of a series of numbers is the reciprocal of the arithmetic mean of the reciprocals of the individual numbers in the series. It is calculated by summating the reciprocals of the numbers, dividing the sum by the number of reciprocals, and then determining the reciprocal of the quotient thus obtained.

Although the adaptability of this mean is somewhat restricted, it can be adapted for certain uses in statistical work. Among these is the averaging of rates of speed. If an aeroplane is flown 120 miles at the rate of 180 miles per hour and 120 miles at the rate of 150 miles per hour the average rate of speed can be calculated as the harmonic mean.

The simple arithmetic mean of the 2 rates of speed at which the plane was flown is 165 miles per hour, obtained by dividing the sum of the 2 rates, 180 and 150 miles per hour, by 2. This mean, 165 miles per hour, may not be the desired measure, however, because it is not actually the correct mean rate of speed at which the plane was flown.

An average of the 2 rates of speed can be calculated by weighting them by the number of minutes that the plane was flown. The flight of 120 miles at the rate of 180 miles per hour required 40 minutes, and the flight of 120 miles at the rate of 150 miles per hour required 48 minutes, so that 240 miles were flown in 98 minutes. By weighting 180 by 40 and 150 by 48, summating the products, and dividing by the sum of the weights (40 plus 48) a mean of 163.6 miles per hour is obtained.

This same result can be obtained by summating the reciprocals of the 2 rates of speed at which the plane was flown (180 and 120 miles per hour), dividing by 2 to obtain the mean of the reciprocals, and then dividing this average of the reciprocals into 1 to determine the reciprocal of the mean of the reciprocals, which is the harmonic

mean. The mean of the reciprocals of 180 and 150 is .00611111, and the reciprocal of .00611111 is 163.6, the harmonic mean. Obviously, the harmonic mean would not be the desired mean in such an instance if the number of miles flown at the 2 rates of speed were different.

Another application of the harmonic mean will be illustrated. Let us assume that a farmer is offering equal quantities of potatoes of different grades at the rate of 2 bushels for a dollar, 4 bushels for a dollar, and 10 bushels for a dollar. These prices are the equivalent of 50 cents per bushel, 25 cents per bushel, and 10 cents per bushel, the arithmetic mean of which is 28-1/3 cents. In terms of bushels, this average of 28-1/3 cents is equivalent to approximately 3.53 bushels per dollar. The harmonic mean of 2, 4, and 10 is also approximately 3.53, obtained as the reciprocal of the arithmetic mean of the reciprocals of 2, 4, and 10. Proceeding on the assumption, therefore, that all quotations made by the farmer are to be given the same weight, the average number of bushels of potatoes that could be purchased for a dollar is approximately 3.53.

This average is not to be confused with the average number of bushels that would be purchased per dollar if the farmer allowed the customer to take 2 bushols of the best grade for 1 dollar, 4 bushels of the next grade for 1 dollar, and 10 bushels of the lowest of the 3 grades for 1 dollar. If 16 bushels were thus purchased in a lot it is obvious that the mean price per bushel of all the potatoes actually purchased for 3 dollars would be 18-3/4 cents, the equivalent of 5-1/3

bushels per dollar. The average of 3.53 bushels per dollar is the number of bushels that could be purchased, as indicated, if each quotation is given equal weight. On this basis of equal weight being given to each quotation the aggregate of 3.53 bushels purchaseable for 1 dollar would be made up of equal quantities of the 3 grades of potatoes.

Geometric Mean

The geometric mean (average) is the nth root of the product of numbers. When there are 2 numbers the geometric mean is the square root of their product, when there are 3 numbers the geometric mean is the cube root of their product, and so on. The geometric mean of 18 and 8, for example, is 12; the geometric mean of 2, 4, and 8 is 4; and the geometric mean of 2, 4, 8, 16, and 32 is 8.

Calculation of the geometric mean is facilitated by the use of logarithms. 2/ To obtain the geometric mean by the use of logarithms it is only necessary to calculate the mean of the logarithms of the numbers and then locate in a logarithmic table the number corresponding to the mean of the logarithms. In illustrating the calculation of the geometric mean by the use of logarithms let us determine the geometric mean of the following:

18 and 8

2, 4, and 8

2, 4, 8, 16, and 32.

2/ Among the authorities on the use of logarithms and on interpolation is Norma L. Goudy of the United States Department of Agriculture.

The logarithms of 18 and 8 are 1.2553 and .9031, respectively, and the mean of their sum is 1.0792. In the logarithmic table it will be found that 1.0792 is the logarithm of 12, which is the geometric mean of 18 and 8, or the square root of their product.

The logarithm of 2 is .3010, the logarithm of 4 is .6021, and the logarithm of 8 is .9031. Summating, we obtain 1.8062, the mean of which is approximately .6021. In the logarithmic table, .6021 will be found to be the logarithm of 4. The geometric mean of 2, 4, and 8, or the cube root of their product, then, is 4.

The logarithms of 2, 4, 8, 16, and 32 are as follows:

.3010
.6021
.9031
1.2041
1.5051

Summating, we obtain 4.5154, the mean of which is approximately .9031, the logarithm of 8. The geometric mean of 2, 4, 8, 16, and 32, therefore, is 8.

This mean has certain uses in the analysis and interpretation of statistical data, among which is its adaptability in the averaging of ratios, such as the ratios of changes in prices from one period to another.

The geometric mean of a series is less than its arithmetic mean and it is greater than its harmonic mean except in those instances in which all the items in the series are of the same magnitude. In such instances the arithmetic, the geometric, and the harmonic means are equal. To illustrate the latter, let us calculate the 3 means of a series of 5 observations, the magnitude of each of which is 12. The

sum of these 10 observations is 60, and the arithmetic mean is 12. Since the logarithm of 12 is 1.0792, the sum of the 5 logarithms is 5.3960; the mean of this sum is 1.0792, which is the logarithm of 12, the geometric mean of the series. The reciprocal of 12 is .083333, the sum of the reciprocals of the series of 5 observations is .416665, and the mean of this sum is .083333. The reciprocal of .08333 is 12, the harmonic mean of the series. 3/

An interesting discussion on the uses of the harmonic and geometric means is contained in chapter 10 of "Statistical Analysis," 1927, by E. E. Day.

3/ The quadratic mean is another mean. It is the root-mean-square, which is determined by extracting the square root of the average of squares of the numbers. For example, the root-mean-square of 2, 8, and 5 is 5.57, the square root of 31, or the square root of the average of the squares of 2, 8, and 5. The arithmetic mean of 2, 8, and 5 is 5.0. The measure of root-mean-square is usually applied only to deviations from a central tendency. Such a measure is commonly called the standard deviation.